P1 1176785

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office

May 27, 2004

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM
THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK
OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT
APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A
FILING DATE.

APPLICATION NUMBER: *60/457,261*
FILING DATE: *March 24, 2003*
RELATED PCT APPLICATION NUMBER: *PCT/US04/07451*

REC'D **0 1 JUN 2004**

| WIPO | PCT |

By Authority of the
COMMISSIONER OF PATENTS AND TRADEMARKS

M. TARVER
Certifying Officer
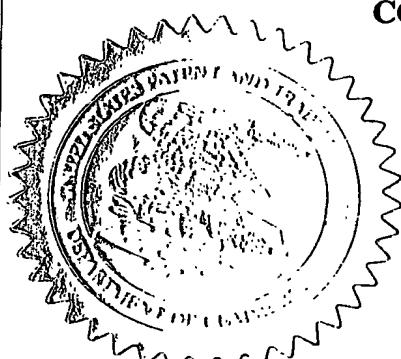
PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

JC950 U.S. PTO
03/24/03
03-26-03
EXPRESS MAIL NO. EV207704954US
PTO/SB/16 (02-01)
Approved for use through 10/31/2002. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
JC996 U.S. PTO
60/457261

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Barbara	Zehentner	Poulsbo, Washington
Dawn	Hayes	Preston, Washington
Raymond L.	Houghton	Bothell, Washington

☐ Additional inventors are being named on the _____ separately numbered sheets attached hereto

TITLE OF THE INVENTION (500 characters max)

METHODS, COMPOSITIONS AND KITS FOR THE DETECTION AND MONITORING OF LUNG CANCER

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☒ Customer Number   —   **32111**

OR   Type Customer Number here

☒ Firm or Individual Name	
Address	
Address	
City		State		ZIP	
Country	US	Telephone	(206) 754-5711	Fax	(206) 754-5994

ENCLOSED APPLICATION PARTS (check all that apply)

☒ Specification Number of Pages **36**   ☐ CD(s), Number _____

☐ Drawing(s) Number of Sheets _____   ☐ Other (specify) _____

☒ Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT

☒ Applicant claims small entity status. See 37 CFR 1.27.

☒ A check or money order for $80 is enclosed to cover the filing fees.

☐ The Commissioner is hereby authorized to charge filing fees to Deposit Account Number: 19-1090

☐ The Commissioner is hereby authorized to charge any deficiency or credit any overpayment to Deposit Account Number: 19-1090

☐ Payment by credit card. Form PTO-2038 is attached.

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____.

Respectfully submitted,			
SIGNATURE	*Richard G. Sharkey*	DATE	March 24, 2003
TYPED or PRINTED NAME	Richard G. Sharkey, Ph.D.	REGISTRATION NO. (if appropriate)	32,629
TELEPHONE	(206) 622-4900	DOCKET NUMBER:	210121.609P1

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

EXPRESS MAIL NO. EV207704954US
PTO/SB/17 (01-03)
Approved for use through 04/30/2003. OMB 0651-0032

FEE TRANSMITTAL
for FY 2003
Patent fees are subject to annual revision.

Complete If Known	
Application Number	
Filing Date	
First Named Inventor	**Barbara Zehentner**
Examiner Name	
Group Art Unit	
Attorney Docket No.	**210121.609P1**

☒ Applicant claims small entity status. See 37 CFR 1.27.

TOTAL AMOUNT OF PAYMENT ($) **80**

METHOD OF PAYMENT

☒ Payment Enclosed:
☒ Check ☐ Credit card ☐ Money Order ☐ Other

☐ Deposit Account:

Deposit Account Number: **19-1090**

Deposit Account Name: **Seed Intellectual Property Law Group PLLC**

The Commissioner is authorized to *(check all that apply)*

☐ Charge fee(s) indicated below   ☒ Credit any overpayments

☐ Charge any additional fee(s) during the pendency of this application

☐ Charge fee(s) indicated below, except for the filing fee

☒ Charge any deficiencies

to the above-identified deposit account.

FEE CALCULATION

1. BASIC FILING FEE

Large Entity		Small Entity			
Fee Code	Fee($)	Fee Code	Fee($)	Fee Description	Fee Paid
1001	750	2001	375	Utility filing fee	
1002	330	2002	165	Design filing fee	
1003	520	2003	260	Plant filing fee	
1004	750	2004	375	Reissue filing fee	
1005	160	2005	80	Provisional filing fee	**80**

SUBTOTAL (1) ($) **80**

2. EXTRA CLAIM FEES

		Extra Claims	Fee from below	Fee Paid
Total Claims		=	*	=
Independent Claims		=	*	=
Multiple Dependent			*	=

Large Entity		Small Entity		
Fee Code	Fee ($)	Fee Code	Fee ($)	Fee Description
1202	18	2202	9	Claims in excess of 20
1201	84	2201	42	Independent claims in excess of 3
1203	280	2203	140	Multiple dependent claim, if not paid
1204	84	2204	42	** Reissue independent claims over original patent
1205	18	2205	9	** Reissue claims in excess of 20 and over original patent

SUBTOTAL (2) ($)

**or number previously paid, if greater; For Reissues, see above

FEE CALCULATION (continued)

3. ADDITIONAL FEES

Large Entity		Small			
Fee Code	Fee ($)	Fee Code	Fee ($)	Fee Description	Fee Paid
1051	130	2051	65	Surcharge - late filing fee or oath	
1052	50	2052	25	Surcharge - late provisional filing fee or cover sheet.	
1053	130	1053	130	Non-English specification	
1812	2520	1812	2520	For filing a request for ex parte reexamination	
1804	920*	1804	920*	Requesting publication of SIR prior to Examiner action	
1805	1840*	1805	1840*	Requesting publication of SIR after Examiner action	
1251	110	2251	55	Extension for reply within first month	
1252	410	2252	205	Extension for reply within second month	
1253	930	2253	465	Extension for reply within third month	
1254	1450	2254	725	Extension for reply within fourth month	
1255	1970	2255	985	Extension for reply within fifth month	
1401	320	2401	160	Notice of Appeal	
1402	320	2402	160	Filing a brief in support of an appeal	
1403	280	2403	140	Request for oral hearing	
1451	1510	1451	1510	Petition to institute a public use proceeding	
1452	110	2452	55	Petition to revive – unavoidable	
1453	1300	2453	650	Petition to revive – unintentional	
1501	1300	2501	650	Utility issue fee (or reissue)	
1502	470	2502	235	Design issue fee	
1503	630	2503	315	Plant issue fee	
1460	130	1460	130	Petitions to the Commissioner	
1807	50	1807	50	Processing fee for provisional applications	
1806	180	1806	180	Submission of Information Disclosure Stmt	
8021	40	8021	40	Recording each patent assignment per property (times number of properties)	
1809	750	2809	375	Filing a submission after final rejection (37 CFR § 1.129(a))	
1810	750	2810	375	For each additional invention to be examined (37 CFR § 1.129(b))	
1801	750	2801	375	Request for Continued Examination (RCE)	
1802	900	1802	900	Request for expedited examination of a design application	

Other fee (specify) _____

*Reduced by Basic Filing Fee Paid

SUBTOTAL (3) ($)

SUBMITTED BY

Name (Print/Type)	**Richard G. Sharkey, Ph.D.**	Registration No. Attorney/Agent	**32,629**
Firm Name/Address			
Signature		Date	**March 24, 2003**
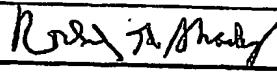
00500
PATENT TRADEMARK OFFICE

367625_1.DOC

APPLICATION DATA SHEET

Application Information

Application number::

Filing Date::

Application Type::   Provisional

Subject Matter::   Utility

Suggested classification::

Suggested Group Art Unit::

CD-ROM or CD-R?::   None

Number of CD disks::

Number of copies of CDs::

Sequence submission?::   Paper

Computer Readable Form (CRF)?::

Number of copies of CRF::

Title ::   METHODS, COMPOSITIONS AND KITS FOR THE DETECTION AND MONITORING OF LUNG CANCER

Attorney Docket Number::   210121.609P1

Request for Early Publication?::   No

Request for Non-Publication?::   No

Suggested Drawing Figure::

Total Drawing Sheets::

Small Entity?::   No

Petition included?::   No

Petition Type::

Licensed U.S. Gov't Agency::

Contract or Grant No::

Secrecy Order in Parent Appl.?::   No

First Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	Germany
Status::	Full Capacity
Given Name::	Barbara
Middle Name::	
Family Name::	Zehentner
Name Suffix::	
City of Residence::	Poulsbo
State or Province of Residence::	WA
Country of Residence::	US
Street of mailing address::	7770 Northeast Harbor View Drive
City of mailing address::	Poulsbo
State or Province of mailing address::	WA
Country of mailing address::	US
Postal or Zip Code of mailing address::	98370

Second Applicant Information

Applicant Authority Type::   Inventor

Primary Citizenship Country::   United States

Status::   Full Capacity

Given Name::   Dawn

Middle Name::

Family Name::   Hayes

Name Suffix::

City of Residence::   Preston

State or Province of Residence::   WA

Country of Residence::   US

Street of mailing address::   8309 308th Avenue Southeast

City of mailing address::   Preston

State or Province of mailing address::   WA

Country of mailing address::   US

Postal or Zip Code of mailing address::   98050

Third Applicant Information

Applicant Authority Type::	Inventor
Primary Citizenship Country::	United States
Status::	Full Capacity
Given Name::	Raymond
Middle Name::	L
Family Name::	Houghton
Name Suffix::	
City of Residence::	Bothell
State or Province of Residence::	WA
Country of Residence::	US
Street of mailing address::	2636 242nd Place Southeast
City of mailing address::	Bothell
State or Province of mailing address::	WA
Country of mailing address::	US
Postal or Zip Code of mailing address::	98021

Correspondence Information

Correspondence Customer Number ::	**32111**
Phone number::	(206) 754-5711
Fax Number:	(206) 754-5994
E-Mail address::	shumate@corixa.com

Initial   3/24/03

Representative Information

Representative Customer Number::		32111

Domestic Priority Information

Application ::	Continuity Type::	Parent Application::	Parent Filing Date::

Foreign Priority Information

Country::	Application number::	Filing Date::	Priority Claimed::

Assignee Information

Assignee name::	Corixa Corporation
Street of mailing address::	1124 Columbia Street, Suite 200
City of mailing address::	Seattle
State or Province of mailing address::	WA
Country of mailing address::	US
Postal or Zip Code of mailing address::	98104

Initial   3/24/03

METHODS, COMPOSITIONS AND KITS
FOR THE DETECTION AND MONITORING OF LUNG CANCER

TECHNICAL FIELD OF THE INVENTION

The present invention relates generally to the field of cancer
5   diagnostics. More specifically, the present invention relates to methods,
compositions and kits for the detection of lung cancer in patients with different
type, stage and grade of tumors that employ oligonucleotide hybridization and/or
amplification to simultaneously detect two or more tissue-specific polynucleotides
in a biological sample suspected of containing lung cancer cells.

10   BACKGROUND OF THE INVENTION

Field of the Invention

Lung cancer remains a significant health problem throughout the
world. The failure of conventional lung cancer treatment regimens can commonly
be attributed, in part, to delayed disease diagnosis. Although significant advances
15   have been made in the area of lung cancer diagnosis, there still remains a need for
improved detection methodologies that permit early, reliable and sensitive
determination of the presence of lung cancer cells.

Description of the Related Art

Lung cancer has the highest mortality rate of any of the cancers and
20   is one of the most difficult to diagnose early. There are an estimated 1 million
deaths annually worldwide for this disease. According to the American Cancer
Society in 2002 alone there were an estimated 169,200 new cases diagnosed and
~ 154,900 deaths. Typically lung cancers are classified into two major types: Non-
Small Cell Lung Carcinomas (NSCLC) comprising squamous, adeno and large cell
25   carcinomas and Small Cell Lung Carcinomas (SCLC). These groups represent

~75% and 25% of all lung tumors respectively with adenocarcinoma and squamous cell carcinoma being the most prevalent forms of NSCLC with large cell carcinomas being ~10%. Within the group of NSCLC, adenocarcinoma is currently the most prominent form of lung cancer in younger persons, women of all.ages,

5 lifetime nonsmokers and long-term former smokers. SCLC typically fall into two subtypes oat cell and intermediate cell. Less common tumors include carcinoid and mesotheliomas among others but these represent only a small percentage of all lung tumors. In almost all cases early diagnosis of NSCLC is elusive and most lung cancers have already metastasized by the time they are detected. Only

10 16.7% are localized on initial diagnosis. If tumors can be detected at a point where they are confined then the combination of chemotherapy and radiation has a possibility of success but.overall the 5year prognosis is very poor with only 10-15% survival rate. The picture with SCLC is even bleaker only 6% localized at initial diagnosis and with 5 year survival rates of ~6%.

15 .   X-ray and computer tomography of the chest and abdomen are frequently used in diagnosis of lung tumors but lack sensitivity for detecting small foci and usually detect tumors that have already metastasized. Sputum cytology as a potential screening method in high-risk individuals has only been partially effective and often does not yield tumor type. To stage the disease CAT scan,

20 MRI or bone scans are used to evaluate the spread of disease. Treatment for lung cancer is typically surgical, radiological or chemotherapy or combinations thereof, but usually with poor outcome due to the late diagnosis of disease.

The current tests for lung cancer lack either the clinical sensitivity to detect early tumors or provide inadequate stage/grade information or lack tumor

25 specificity due to their originating from other tumor types or being present in benign lung disorders. There is therefore a need to develop specific tests that can improve lung cancer diagnosis and prognosis and potentially differentiate between NSCLC and SCLC. The present invention achieves these and other related objectives by providing methods that are useful for the identification of tissue-

specific polynucleotides, in particular tumor-specific polynucleotides, as well as antibodies and methods, compositions and kits for the detection and monitoring of cancer cells in a patient afflicted with the disease.

BRIEF SUMMARY OF THE INVENTION

5   The present invention provides methods for detecting the presence of lung cancer cells in a patient. Such methods comprise the steps of: (a) obtaining a biological sample from the patient; (b) contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs

10   hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof (c) amplifying the polynucleotides; and (d) detecting the amplified polynucleotides; wherein the presence of one or more of the amplified polynucleotides indicates the presence of lung cancer cells in the patient.

15   By some embodiments, detection of the amplified polynucleotides may be preceded by a fractionation step such as, for example, gel electrophoresis. Alternatively or additionally, detection of the amplified polynucleotides may be achieved by hybridization of a labeled oligonucleotide probe that hybridizes specifically, under moderately stringent conditions, to such polynucleotides.

20   Oligonucleotide labeling may be achieved by incorporating a radiolabeled nucleotide or by incorporating a fluorescent label.

In certain preferred embodiments, cells of a specific tissue type may be enriched from the biological sample prior to the steps of detection. Enrichment may be achieved by a methodology selected from the group consisting of cell

25   capture and cell depletion. Exemplary cell capture methods include immunocapture and comprise the steps of: (a) adsorbing an antibody to a tissue-specific cell surface to cells said biological sample; (b) separating the antibody adsorbed tissue-specific cells from the remainder of the biological sample.

Exemplary cell depletion may be achieved by cross-linking red cells and white cells followed by a subsequent fractionation step to remove the cross-linked cells.xxx

Alternative embodiments of the present invention provide methods for determining the presence or absence of lung cancer in a patient, comprising

5 the steps of: (a) contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins; (b) detecting in the sample a level of at least one of the polynucleotides (such as, for example, mRNA) that hybridize to the oligonucleotides; and (c) comparing the level of polynucleotides that hybridize to

10 the oligonucleotides with a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer in the patient. Within certain embodiments, the amount of mRNA is detected via polymerase chain reaction using, for example, at least one oligonucleotide primer that hybridizes to a polynucleotide encoding a polypeptide as recited above, or a complement of such

15 a polynucleotide. Within other embodiments, the amount of mRNA is detected using a hybridization technique, employing an oligonucleotide probe that hybridizes to a polynucleotide that encodes a polypeptide as recited above, or a complement of such a polynucleotide.

In related aspects, methods are provided for monitoring the

20 progression of lung cancer in a patient, comprising the steps of: (a) contacting a biological sample obtained from a patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode lung tumor proteins; (b) detecting in the sample an amount of the polynucleotides that hybridize to the oligonucleotides; (c) repeating steps (a) and (b) using a biological sample obtained

25 from the patient at a subsequent point in time; and (d) comparing the amount of polynucleotide detected in step (c) with the amount detected in step (b) and therefrom monitoring the progression of the cancer in the patient.

Certain embodiments of the present invention provide that the step of amplifying said first polynucleotide and said second polynucleotide is achieved by the polymerase chain reaction (PCR).

The present invention also provides kits that are suitable for
5   performing the detection methods of the present invention. Exemplary kits comprise oligonucleotide primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Within certain embodiments, kits according to the present invention may also comprise a nucleic acid polymerase and suitable buffer.

10   These and other aspects of the present invention will become apparent upon reference to the following detailed description and attached drawings. All references disclosed herein are hereby incorporated by reference in their entirety as if each was incorporated individually.

BRIEF DESCRIPTION OF SEQUENCE IDENTIFIERS

15   SEQ ID NO: 1 is the determined cDNA sequence L762P.

SEQ ID NO: 2 is the amino acid sequence encoded by the sequence of SEQ ID NO: 1.

SEQ ID NO: 3 is the determined cDNA sequence L984P.

SEQ ID NO: 4 is the amino acid sequence encoded by the sequence
20   of SEQ ID NO: 3.

SEQ ID NO: 5 is the determined cDNA sequence 550S.

SEQ ID NO: 6 is the amino acid sequence encoded by the sequence of SEQ ID NO: 5.

SEQ ID NO: 7 is the determined cDNA sequence L552S.

25   SEQ ID NO: 8 is the amino acid sequence encoded by the sequence of SEQ ID NO: 7.

SEQ ID NO:9 is the DNA sequence of L552S INT forward primer.

SEQ ID NO:10 is the DNA sequence of L552S INT reverse primer.

SEQ ID NO:11 is the DNA sequence of L552S Taqman probe.

SEQ ID NO:12 is the DNA sequence of L550S INT forward primer.

SEQ ID NO:13 is the DNA sequence of L550S INT reverse primer.

SEQ ID NO:14 is the DNA sequence of L550S Taqman probe.

5   SEQ ID NO:15 is the DNA sequence of L726P INT forward primer.

SEQ ID NO:16 is the DNA sequence of L726P INT reverse primer.

SEQ ID NO:17 is the DNA sequence of L726P Taqman probe.

SEQ ID NO:18 is the DNA sequence of L984P INT forward primer.

SEQ ID NO:19 is the DNA sequence of L984P INT reverse primer.

10   SEQ ID NO:20 is the DNA sequence of L984P Taqman probe.

DETAILED DESCRIPTION OF THE INVENTION

As noted above, the present invention is directed generally to methods that are suitable for the identification of tissue-specific polynucleotides as well as to methods, compositions and kits that are suitable for the diagnosis and

15   monitoring of lung cancer, in particular such methods, compositions and kits are suitable for use in the diagnosis, differentiation and/or prognosis of NSCLC and SCLC.   Such diagnostic methods will form the basis for a molecular diagnostic test for detecting lung cancer metastases in lung tissue and for the detection of anchorage independent lung cancer cells in blood as well as in mediastinal lymph

20   nodes of distant metastases.

A variety of genes have been identified as over-expressed in lung tumors, in particular squamous or adeno forms of NSCLC or small cell carcinomas. These include, but are not limited to: L762P, L984P, L550S/L548S, L552S/L547S, L552/L547S, L200T, L514S, L551S, L587S, L763S, L773P, L801P. L985P,

25   L1058C, L1081C, L523S, OF1783P, B307D (WIPO International Patent Application Nos: WO 99/47674, published September 23, 1999; WO 00/61612, published October 19, 2000; WO 02/00174, published January 3, 2002; WO 02/47534, published June 20, 2002; WO 01/72295, published October 4, 2001;

6

WO 02/092001, published November 21, 2002; WO 01/00828, published January 1, 2001; WO 02/04514, published January 17, 2002; WO 01/92525, published December 6, 2002; WO 02/02623, published January 10, 2002. US Patent Nos: Wang et al., 6,482,597, issued November 22, 2002; Wang et al., 6,518,256, issued
5 February 11, 2003; Wang et al., 6,426,072, issued July 30, 2002; Reed et al., 6,210,883, issued April 3, 2001; Wang et al., 6,504,010, issued January 7, 2003; Wang et al., 6,509,448, issued January 21, 2003. Wang *et al; Oncogene*; *21*(49):7598-604, 2002 (collagen type XI alpha 1).).

These genes were identified and characterized using PCR and cDNA
10 library subtractions as well as electronic subtractions with each of the tumor types individually. The cDNAs identified were then evaluated by microarray then by Real Time PCR on tissue panels to identify specific expression patterns. Table 1 highlights the specificity of these genes for either adeno or squamous forms of NSCLC or both as well as genes specific for small cell lung carcinomas. In some
15 cases reactivity with large cell carcinomas has also been identified by Real Time PCR analysis.

TABLE 1

Gene	Squamous	Adeno	Small cell	Large cell	Normal Lung
L762P	+++++	+			-
L984P		+	+++		-
L550S/L548S		+++++	+		-
L552S/L547S	++	+++++			-
L200T	+	++		++	-
L514S	++++	++++			-
L551S		++++		+/-	-
L587S	+	+	+++	+	-
L763P	+++++				-
L773P	+++	+++			-
L801P	++++	++++		++	-
L978P	+	++	+++++	+/-	-
L985P		+	+++++		-
L1058C			++		-

L1081C			++		-
L523S	+++++	+++++	+	++	-
OF 1783P			+++++		-
B307D	++	++		+	-

Identification of Tissue-specific Polynucleotides

Certain embodiments of the present invention provide methods, compositions and kits for the detection of lung cancer cells within a biological

5   sample from patients with different type, stage and grade of tumors. These methods comprise the step of detecting one or more tissue-specific polynucleotide(s) from a patient's biological sample the over-expression of which polynucleotides indicates the presence of lung cancer cells within the patient's biological sample. Accordingly, the present invention also provides methods that

10   are suitable for the identification of tissue-specific polynucleotides. As used herein, the phrases "tissue-specific polynucleotides" or "tumor-specific polynucleotides" are meant to include all polynucleotides that are at least two-fold over-expressed as compared to one or more control tissues. As discussed in further detail herein below, over-expression of a given polynucleotide may be

15   assessed, for example, by microarray and/or quantitative real-time polymerase chain reaction (Real-time PCR^{TM}) methodologies.

Exemplary methods for detecting tissue-specific polynucleotides may comprise the steps of: (a) performing a genetic subtraction to identify a pool of polynucleotides from a tissue of interest; (b) performing a DNA microarray analysis

20   to identify a first subset of said pool of polynucleotides of interest wherein each member polynucleotide of said first subset is at least two-fold over-expressed in said tissue of interest as compared to a control tissue; and (c) performing a quantitative polymerase chain reaction analysis on polynucleotides within said first subset to identify a second subset of polynucleotides that are at least two-fold

25   over-expressed as compared to said control tissue.

8

Polynucleotides Generally

As used herein, the term "polynucleotide" refers generally to either DNA or RNA molecules. Polynucleotides may be naturally occurring as normally found in a biological sample such as blood, serum, lymph node, bone marrow,

5   sputum, urine and tumor biopsy samples. Alternatively, polynucleotides may be derived synthetically by, for example, a nucleic acid polymerization reaction. As will be recognized by the skilled artisan, polynucleotides may be single-stranded (coding or antisense) or double-stranded, and may be DNA (genomic, cDNA or synthetic) or RNA molecules. RNA molecules include HnRNA molecules, which

10   contain introns and correspond to a DNA molecule in a one-to-one manner, and mRNA molecules, which do not contain introns. Additional coding or non-coding sequences may, but need not, be present within a polynucleotide of the present invention, and a polynucleotide may, but need not, be linked to other molecules and/or support materials.

15   Polynucleotides may comprise a native sequence (*i.e.,* an endogenous sequence that encodes a tumor protein, such as a lung tumor protein, or a portion thereof) or may comprise a variant, or a biological or antigenic functional equivalent of such a sequence. Polynucleotide variants may contain one or more substitutions, additions, deletions and/or insertions, as further

20   described below. The term "variants" also encompasses homologous genes of xenogenic origin.

When comparing polynucleotide or polypeptide sequences, two sequences are said to be "identical" if the sequence of nucleotides or amino acids in the two sequences is the same when aligned for maximum correspondence, as

25   described below. Comparisons between two sequences are typically performed by comparing the sequences over a comparison window to identify and compare local regions of sequence similarity. A "comparison window" as used herein, refers to a segment of at least about 20 contiguous positions, usually 30 to about 75, 40 to about 50, in which a sequence may be compared to a reference sequence of the

9

same number of contiguous positions after the two sequences are optimally aligned.

Optimal alignment of sequences for comparison may be conducted using the Megalign program in the Lasergene suite of bioinformatics software

5   (DNASTAR, Inc., Madison, WI), using default parameters. This program embodies several alignment schemes described in the following references: Dayhoff, M.O. (1978) A model of evolutionary change in proteins – Matrices for detecting distant relationships. In Dayhoff, M.O. (ed.) Atlas of Protein Sequence and Structure, National Biomedical Research Foundation, Washington DC Vol. 5, Suppl. 3, pp.

10   345-358; Hein J. (1990) Unified Approach to Alignment and Phylogenes pp. 626-645 *Methods in Enzymology* vol. 183, Academic Press, Inc., San Diego, CA; Higgins, D.G. and Sharp, P.M. (1989) *CABIOS* 5:151-153; Myers, E.W. and Muller W. (1988) *CABIOS* 4:11-17; Robinson, E.D. (1971) *Comb. Theor* 11:105; Santou, N. Nes, M. (1987) *Mol. Biol. Evol.* 4:406-425; Sneath, P.H.A. and Sokal, R.R.

15   (1973) *Numerical Taxonomy – the Principles and Practice of Numerical Taxonomy*, Freeman Press, San Francisco, CA; Wilbur, W.J. and Lipman, D.J. (1983) *Proc. Natl. Acad., Sci. USA* 80:726-730.

Alternatively, optimal alignment of sequences for comparison may be conducted by the local identity algorithm of Smith and Waterman (1981) *Add. APL.*

20   *Math* 2:482, by the identity alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity methods of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. USA* 85: 2444, by computerized implementations of these algorithms (GAP, BESTFIT, BLAST, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group (GCG), 575 Science Dr.,

25   Madison, WI), or by inspection.

One preferred example of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1977) *Nucl. Acids Res.* 25:3389-3402 and Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410, respectively. BLAST and

BLAST 2.0 can be used, for example with the parameters described herein, to determine percent sequence identity for the polynucleotides and polypeptides of the invention. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information. In one illustrative

5 example, cumulative scores can be calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always >0) and N (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix can be used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls

10 off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of

15 11, and expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff and Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915) alignments, (B) of 50, expectation (E) of 10, M=5, N=-4 and a comparison of both strands.

Preferably, the "percentage of sequence identity" is determined by comparing two optimally aligned sequences over a window of comparison of at

20 least 20 positions, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less, usually 5 to 15 percent, or 10 to 12 percent, as compared to the reference sequences (which does not comprise additions or deletions) for optimal alignment of the two sequences. The percentage is

25 calculated by determining the number of positions at which the identical nucleic acid bases or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the reference sequence (*i.e.*, the window size) and multiplying the results by 100 to yield the percentage of sequence identity.

11

Therefore, the present invention encompasses polynucleotide and polypeptide sequences having substantial identity to the sequences disclosed herein, for example those comprising at least 50% sequence identity, preferably at least 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99%

5   or higher, sequence identity compared to a polynucleotide or polypeptide sequence of this invention using the methods described herein, (e.g., BLAST analysis using standard parameters, as described below). One skilled in this art will recognize that these values can be appropriately adjusted to determine corresponding identity of proteins encoded by two nucleotide sequences by taking

10   into account codon degeneracy, amino acid similarity, reading frame positioning and the like.

In additional embodiments, the present invention provides isolated polynucleotides and polypeptides comprising various lengths of contiguous stretches of sequence identical to or complementary to one or more of the

15   sequences disclosed herein. For example, polynucleotides are provided by this invention that comprise at least about 15, 20, 30, 40, 50, 75, 100, 150, 200, 300, 400, 500 or 1000 or more contiguous nucleotides of one or more of the sequences disclosed herein as well as all intermediate lengths there between. It will be readily understood that "intermediate lengths", in this context, means any length

20   between the quoted values, such as 16, 17, 18, 19, etc.; 21, 22, 23, etc.; 30, 31, 32, etc.; 50, 51, 52, 53, etc.; 100, 101, 102, 103, etc.; 150, 151, 152, 153, etc.; including all integers through 200-500; 500-1,000, and the like.

The polynucleotides of the present invention, or fragments thereof, regardless of the length of the coding sequence itself, may be combined with other

25   DNA sequences, such as promoters, polyadenylation signals, additional restriction enzyme sites, multiple cloning sites, other coding segments, and the like, such that their overall length may vary considerably. It is therefore contemplated that a nucleic acid fragment of almost any length may be employed, with the total length preferably being limited by the ease of preparation and use in the intended

recombinant DNA protocol. For example, illustrative DNA segments with total lengths of about 10,000, about 5000, about 3000, about 2,000, about 1,000, about 500, about 200, about 100, about 50 base pairs in length, and the like, (including all intermediate lengths) are contemplated to be useful in many implementations of

5   this invention.

In other embodiments, the present invention is directed to polynucleotides that are capable of hybridizing under moderately stringent conditions to a polynucleotide sequence provided herein, or a fragment thereof, or a complementary sequence thereof. Hybridization techniques are well known in

10   the art of molecular biology. For purposes of illustration, suitable moderately stringent conditions for testing the hybridization of a polynucleotide of this invention with other polynucleotides include prewashing in a solution of 5 X SSC, 0.5% SDS, 1.0 mM EDTA (pH 8.0); hybridizing at 50°C-65°C, 5 X SSC, overnight; followed by washing twice at 65°C for 20 minutes with each of 2X, 0.5X and 0.2X SSC

15   containing 0.1% SDS.

Moreover, it will be appreciated by those of ordinary skill in the art that, as a result of the degeneracy of the genetic code, there are many nucleotide sequences that encode a polypeptide as described herein. Some of these polynucleotides bear minimal homology to the nucleotide sequence of any native

20   gene. Nonetheless, polynucleotides that vary due to differences in codon usage are specifically contemplated by the present invention. Further, alleles of the genes comprising the polynucleotide sequences provided herein are within the scope of the present invention. Alleles are endogenous genes that are altered as a result of one or more mutations, such as deletions, additions and/or substitutions

25   of nucleotides. The resulting mRNA and protein may, but need not, have an altered structure or function. Alleles may be identified using standard techniques (such as hybridization, amplification and/or database sequence comparison).

13

Microarray Analyses

Polynucleotides that are suitable for detection according to the methods of the present invention may be identified, as described in more detail below, by screening a microarray of cDNAs for tissue and/or tumor-associated

5   expression (e.g., expression that is at least two-fold greater in a tumor than in normal tissue, as determined using a representative assay provided herein). Such screens may be performed, for example, using a Synteni microarray (Palo Alto, CA) according to the manufacturer's instructions (and essentially as described by Schena et al., Proc. Natl. Acad. Sci. USA 93:10614-10619 (1996) and Heller et al.,

10   Proc. Natl. Acad. Sci. USA 94:2150-2155 (1997)).

Microarray is an effective method for evaluating large numbers of genes but due to its limited sensitivity it may not accurately determine the absolute tissue distribution of low abundance genes or may underestimate the degree of overexpression of more abundant genes due to signal saturation. For those genes

15   showing overexpression by microarray expression profiling, further analysis was performed using quantitative RT-PCR based on Taqman™ probe detection, which comprises a greater dynamic range of sensitivity. Several different panels of normal and tumor tissues, distant metastases and cell lines were used for this purpose.

20   ## Quantitative Real-time Polymerase Chain Reaction

Suitable polynucleotides according to the present invention may be further characterized or, alternatively, originally identified by employing a quantitative PCR methodology such as, for example, the Real-time PCR methodology. By this methodology, tissue and/or tumor samples, such as, e.g.,

25   metastatic tumor samples, may be tested along side the corresponding normal tissue sample and/or a panel of unrelated normal tissue samples.

Real-time PCR (see Gibson et al., Genome Research 6:995-1001, 1996; Heid et al., Genome Research 6:986-994, 1996) is a technique that

evaluates the level of PCR product accumulation during amplification. This technique permits quantitative evaluation of mRNA levels in multiple samples. Briefly, mRNA is extracted from tumor and normal tissue and cDNA is prepared using standard techniques.

5   Real-time PCR may, for example, be performed either on the ABI 7700 Prism or on a GeneAmp® 5700 sequence detection system (PE Biosystems, Foster City, CA). The 7700 system uses a forward and a reverse primer in combination with a specific probe with a 5' fluorescent reporter dye at one end and a 3' quencher dye at the other end (Taqman™). When the Real-time PCR is

10   performed using Taq DNA polymerase with 5'-3' nuclease activity, the probe is cleaved and begins to fluoresce allowing the reaction to be monitored by the increase in fluorescence (Real-time). The 5700 system uses SYBR® green, a fluorescent dye, that only binds to double stranded DNA, and the same forward and reverse primers as the 7700 instrument. Matching primers and fluorescent

15   probes may be designed according to the primer express program (PE Biosystems, Foster City, CA). Optimal concentrations of primers and probes are initially determined by those of ordinary skill in the art. Control (e.g., β-actin) primers and probes may be obtained commercially from, for example, Perkin Elmer/Applied Biosystems (Foster City, CA).

20   To quantitate the amount of specific RNA in a sample, a standard curve is generated using a plasmid containing the gene of interest. Standard curves are generated using the Ct values determined in the real-time PCR, which are related to the initial cDNA concentration used in the assay. Standard dilutions ranging from 10-10^6 copies of the gene of interest are generally sufficient. In

25   addition, a standard curve is generated for the control sequence. This permits standardization of initial RNA content of a tissue sample to the amount of control for comparison purposes.

          In accordance with the above, and as described further below, the present invention provides the illustrative lung tissue- and/or tumor-specific

15

polynucleotides L552S, L550S, L762P and L984P having sequences set forth in SEQ ID NOs: 1, 3, 5 and 7, illustrative polypeptides encoded thereby having amino acid sequences set forth in SEQ ID NO: 2, 4, 6 and 8 that may be suitably employed in the detection of cancer, more specifically, lung cancer.

5 Methodologies for the Detection of Cancer

In general, a cancer cell may be detected in a patient based on the presence of one or more polynucleotides within cells of a biological sample (for example, blood, lymph nodes, bone marrow, sera, sputum, urine and/or tumor biopsies) obtained from the patient. In other words, such polynucleotides may be
10 used as markers to indicate the presence or absence of a cancer such as, e.g., lung cancer.

As discussed in further detail herein, the present invention achieves these and other related objectives by providing a methodology for the simultaneous detection of more than one polynucleotide, the presence of which is
15 diagnostic of the presence of lung cancer cells in a biological sample. Each of the various cancer detection methodologies disclosed herein have in common a step of hybridizing one or more oligonucleotide primers and/or probes, the hybridization of which is demonstrative of the presence of a tumor- and/or tissue-specific polynucleotide. Depending on the precise application contemplated, it may be
20 preferred to employ one or more intron-spanning oligonucleotides that are inoperative against polynucleotide of genomic DNA and, thus, these oligonucleotides are effective in substantially reducing and/or eliminating the detection of genomic DNA in the biological sample.

Further disclosed herein are methods for enhancing the sensitivity of
25 these detection methodologies by subjecting the biological samples to be tested to one or more cell capture and/or cell depletion methodologies.

By certain embodiments of the present invention, the presence of lung cancer cell in a patient may be determined by employing the following steps:

16

(a) contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins as described herein; (b) detecting in the sample a level of at least one of the polynucleotides (such as, for example, mRNA) that hybridize

5   to the oligonucleotides; and (c) comparing the level of polynucleotides that hybridize to the oligonucleotides with a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer in the patient.

To permit hybridization under assay conditions, oligonucleotide primers and probes should comprise an oligonucleotide sequence that has at least

10   about 60%, preferably at least about 75% and more preferably at least about 90%, identity to a portion of a polynucleotide encoding a lung tumor protein that is at least 10 nucleotides, and preferably at least 20 nucleotides, in length. Preferably, oligonucleotide primers hybridize to a polynucleotide encoding a polypeptide described herein under moderately stringent conditions, as defined above.

15   Oligonucleotide primers which may be usefully employed in the diagnostic methods described herein preferably are at least 10-40 nucleotides in length. In a preferred embodiment, the oligonucleotide primers comprise at least 10 contiguous nucleotides, more preferably at least 15 contiguous nucleotides, of a DNA molecule having a sequence recited in SEQ ID NO: 1, 3, 5 or 7. Techniques for

20   both PCR based assays and hybridization assays are well known in the art (see, for example, Mullis et al., Cold Spring Harbor Symp. Quant. Biol., 51:263, 1987; Erlich ed., PCR Technology, Stockton Press, NY, 1989).

The present invention also provides amplification-based methods for detecting the presence of lung cancer cells in a patient. Exemplary methods

25   comprise the steps of (a) obtaining a biological sample from the patient; (b) contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof (c) amplifying the

17

polynucleotides; and (d) detecting the amplified polynucleotides; wherein the presence of one or more of the amplified polynucleotides indicates the presence of lung cancer cells in the patient.

Methods according to the present invention are suitable for

5   identifying polynucleotides obtained from a wide variety of biological sample such as, for example, blood, serum, lymph node, bone marrow, sputum, urine and tumor biopsy sample, among others.

Certain exemplary embodiments of the present invention provide methods wherein the polynucleotides to be detected are selected from the group

10   consisting of L762, L984, L550, and L552. Alternatively and/or additionally, polynucleotides to be detected may be selected from the group consisting of those depicted in SEQ ID NOs: 1, 3, 5, and 7.

Suitable exemplary oligonucleotide probes and/or primers that may be used according to the methods of the present invention are disclosed herein.

15   In certain preferred embodiments that eliminate the background detection of genomic DNA, the oligonucleotides may be intron spanning oligonucleotides.

Depending on the precise application contemplated, the artisan may prefer to detect the tissue- and/or tumor-specific polynucleotides by detecting a radiolabel and detecting a fluorophore. More specifically, the oligonucleotide

20   probe and/or primer may comprises a detectable moiety such as, for example, a radiolabel and/or a fluorophore.

Alternatively or additionally, methods of the present invention may also comprise a step of fractionation prior to detection of the tissue- and/or tumor-specific polynucleotides such as, for example, by gel electrophoresis.

25   In other embodiments, methods described herein may be used as to monitor the progression of cancer. By these embodiments, assays as provided for the diagnosis of lung cancer may be performed over time, and the change in the level of reactive polypeptide(s) or polynucleotide(s) evaluated. For example, the assays may be performed every 24-72 hours for a period of 6 months to 1 year,

and thereafter performed as needed. In general, a cancer is progressing in those patients in whom the level of polypeptide or polynucleotide detected increases over time. In contrast, the cancer is not progressing when the level of reactive polypeptide or polynucleotide either remains constant or decreases with time.

5   Certain *in vivo* diagnostic assays may be performed directly on a tumor. One such assay involves contacting tumor cells with a binding agent. The bound binding agent may then be detected directly or indirectly via a reporter group. Such binding agents may also be used in histological applications. Alternatively, polynucleotide probes may be used within such applications.

10   As noted above, to improve sensitivity, multiple lung tumor protein markers may be assayed within a given sample. It will be apparent that binding agents specific for different proteins provided herein may be combined within a single assay. Further, multiple primers or probes may be used concurrently. The selection of tumor protein markers may be based on routine experiments to

15   determine combinations that results in optimal sensitivity. In addition, or alternatively, assays for tumor proteins provided herein may be combined with assays for other known tumor antigens.

Cell Enrichment

          In other aspects of the present invention, cell capture technologies

20   may be used prior to polynucleotide detection to improve the sensitivity of the various detection methodologies disclosed herein.

          Exemplary cell enrichment methodologies employ immunomagnetic beads that are coated with specific monoclonal antibodies to surface cell markers, or tetrameric antibody complexes, may be used to first enrich or positively select

25   cancer cells in a sample. Various commercially available kits may be used, including Dynabeads® Epithelial Enrich (Dynal Biotech, Oslo, Norway), StemSep™ (StemCell Technologies, Inc., Vancouver, BC), and RosetteSep (StemCell Technologies). The skilled artisan will recognize that other readily

available methodologies and kits may also be suitably employed to enrich or positively select desired cell populations.

Dynabeads® Epithelial Enrich contains magnetic beads coated with mAbs specific for two glycoprotein membrane antigens expressed on normal and neoplastic epithelial tissues. The coated beads may be added to a sample and the sample then applied to a magnet, thereby capturing the cells bound to the beads. The unwanted cells are washed away and the magnetically isolated cells eluted from the beads and used in further analyses.

RosetteSep can be used to enrich cells directly from a blood sample and consists of a cocktail of tetrameric antibodies that target a variety of unwanted cells and crosslinks them to glycophorin A on red blood cells (RBC) present in the sample, forming rosettes. When centrifuged over Ficoll, targeted cells pellet along with the free RBC.

The combination of antibodies in the depletion cocktail determines which cells will be removed and consequently which cells will be recovered. Antibodies that are available include, but are not limited to: CD2, CD3, CD4, CD5, CD8, CD10, CD11b, CD14, CD15, CD16, CD19, CD20, CD24, CD25, CD29, CD33, CD34, CD36, CD38, CD41, CD45, CD45RA, CD45RO, CD56, CD66B, CD66e, HLA-DR, IgE, and TCRαβ. Additionally, it is contemplated in the present invention that mAbs specific for lung tumor antigens, can be developed and used in a similar manner. For example, mAbs that bind to tumor-specific cell surface antigens may be conjugated to magnetic beads, or formulated in a tetrameric antibody complex, and used to enrich or positively select metastatic lung tumor cells from a sample. Such a system can be used to evaluate blood samples from different forms of lung cancers, in particular adneo and squamous forms of NSCLC and small cell carcinomas for the presence of circulating tumor cells using the inventive multiplex PCR assay as described herein.

Once a sample is enriched or positively selected, cells may be further analyzed. For example, the cells may be lysed and RNA isolated. RNA may then

20

be subjected to RT-PCR analysis using lung tumor-specific multiplex primers in a Real-time PCR assay as described herein.

In another aspect of the present invention, cell capture technologies may be used in conjunction with Real-Time PCR to provide a more sensitive tool for detection of metastatic cells expressing lung tumor antigens.

Yet another method that may be employed is an anti-ganglioside G_{M1}/G_{M1} cell capture antibody system. Gangliosides are cell membrane bound glycosphingolipids, several species of which have been shown to be over-expressed on the cell surface of most cancers of neuroectodermal and epithelial origin, in particular lung cancer. Cell surface expression of G_{M2} is seen in several types of lung cancer, particularly in SCLC which make it an attractive target for a monoclonal antibody based lung cancer immunotherapy and also for use as a capture method in conjunction with G_{M1}.

Probes and Primers

As noted above and as described in further detail herein, certain methods, compositions and kits according to the present invention utilize two or more oligonucleotide primer pairs for the detection of lung cancer. The ability of such nucleic acid probes to specifically hybridize to a sequence of interest will enable them to be of use in detecting the presence of complementary sequences in a biological sample.

Alternatively, in other embodiments, the probes and/or primers of the present invention may be employed for detection via nucleic acid hybridization. As such, it is contemplated that nucleic acid segments that comprise a sequence region of at least about 15 nucleotide long contiguous sequence that has the same sequence as, or is complementary to, a 15 nucleotide long contiguous sequence of a polynucleotide to be detected will find particular utility. Longer contiguous identical or complementary sequences, e.g., those of about 20, 30, 40, 50, 100,

200, 500, 1000 (including all intermediate lengths) and even up to full length sequences will also be of use in certain embodiments.

Oligonucleotide primers having sequence regions consisting of contiguous nucleotide stretches of 10-14, 15-20, 30, 50, or even of 100-200 nucleotides or so (including intermediate lengths as well), identical or complementary to a polynucleotide to be detected, are particularly contemplated as primers for use in amplification reactions such as, e.g., the polymerase chain reaction (PCR™). This would allow a polynucleotide to be analyzed, both in diverse biological samples such as, for example, blood, lymph nodes and bone marrow.

The use of a primer of about 15-25 nucleotides in length allows the formation of a duplex molecule that is both stable and selective. Molecules having contiguous complementary sequences over stretches greater than 15 bases in length are generally preferred, though, in order to increase stability and selectivity of the hybrid, and thereby improve the quality and degree of specific hybrid molecules obtained. One will generally prefer to design primers having gene-complementary stretches of 15 to 25 contiguous nucleotides, or even longer where desired.

Primers may be selected from any portion of the polynucleotide to be detected. All that is required is to review the sequence, such as those exemplary polynucleotides set forth herein or to any continuous portion of the sequence, from about 15-25 nucleotides in length up to and including the full length sequence, that one wishes to utilize as a primer. The choice of primer sequences may be governed by various factors. For example, one may wish to employ primers from towards the termini of the total sequence. The exemplary primers disclosed herein may optionally be used for their ability to selectively form duplex molecules with complementary stretches of the entire polynucleotide of interest such as those set forth SEQ ID NOs: 1, 3 ,5 or 7.

22

The present invention further provides the nucleotide sequence of various exemplary oligonucleotide primers and probes, that may be used, as described in further detail herein, according to the methods of the present invention for the detection of cancer.

5   Oligonucleotide primers according to the present invention may be readily prepared routinely by methods commonly available to the skilled artisan including, for example, directly synthesizing the primers by chemical means, as is commonly practiced using an automated oligonucleotide synthesizer. Depending on the application envisioned, one will typically desire to employ varying conditions

10 of hybridization to achieve varying degrees of selectivity of probe towards target sequence. For applications requiring high selectivity, one will typically desire to employ relatively stringent conditions to form the hybrids, e.g., one will select relatively low salt and/or high temperature conditions, such as provided by a salt concentration of from about 0.02 M to about 0.15 M salt at temperatures of from

15 about 50°C to about 70°C. Such selective conditions tolerate little, if any, mismatch between the probe and the template or target strand, and would be particularly suitable for isolating related sequences.

Polynucleotide Amplification Techniques

          Each of the specific embodiments outlined herein for the detection of

20 lung cancer has in common the detection of a tissue- and/or tumor-specific polynucleotide via the hybridization of one or more oligonucleotide primers and/or probes. Depending on such factors as the relative number of cancer cells present in the biological sample and/or the level of polynucleotide expression within each lung cancer cell, it may be preferred to perform an amplification step prior to

25 performing the steps of detection. For example, at least two oligonucleotide primers may be employed in a polymerase chain reaction (PCR) based assay to amplify a portion of a lung tumor cDNA derived from a biological sample, wherein at least one of the oligonucleotide primers is specific for (i.e., hybridizes to) a

23

polynucleotide encoding the lung tumor protein. The amplified cDNA may optionally be subjected to a fractionation step such as, for example, gel electrophoresis.

A number of template dependent processes are available to amplify
5   the target sequences of interest present in a sample. One of the best known amplification methods is the polymerase chain reaction (PCR™) which is described in detail in U.S. Patent Nos. 4,683,195, 4,683,202 and 4,800,159. Briefly, in PCR™, two primer sequences are prepared which are complementary to regions on opposite complementary strands of the target sequence. An excess of
10   deoxynucleoside triphosphates is added to a reaction mixture along with a DNA polymerase (e.g., Taq polymerase). If the target sequence is present in a sample, the primers will bind to the target and the polymerase will cause the primers to be extended along the target sequence by adding on nucleotides. By raising and lowering the temperature of the reaction mixture, the extended primers will
15   dissociate from the target to form reaction products, excess primers will bind to the target and to the reaction product and the process is repeated. Preferably reverse transcription and PCR™ amplification procedure may be performed in order to quantify the amount of mRNA amplified. Polymerase chain reaction methodologies are well known in the art.
20   One preferred methodology for polynucleotide amplification employs RT-PCR, in which PCR is applied in conjunction with reverse transcription. Typically, RNA is extracted from a biological sample, such as blood, serum, lymph node, bone marrow, sputum, urine and tumor biopsy samples, and is reverse transcribed to produce cDNA molecules. PCR amplification using at least one
25   specific primer generates a cDNA molecule, which may be separated and visualized using, for example, gel electrophoresis. Amplification may be performed on biological samples taken from a patient and from an individual who is not afflicted with a cancer. The amplification reaction may be performed on several dilutions of cDNA spanning two orders of magnitude. A two-fold or greater

increase in expression in several dilutions of the test patient sample as compared to the same dilutions of the non-cancerous sample is typically considered positive.

Any of a variety of commercially available kits may be used to perform the amplification step. One such amplification technique is inverse PCR

5   (see Triglia et al., Nucl. Acids Res. 16:8186, 1988), which uses restriction enzymes to generate a fragment in the known region of the gene. The fragment is then circularized by intramolecular ligation and used as a template for PCR with divergent primers derived from the known region. Within an alternative approach, sequences adjacent to a partial sequence may be retrieved by amplification with a

10   primer to a linker sequence and a primer specific to a known region. The amplified sequences are typically subjected to a second round of amplification with the same linker primer and a second primer specific to the known region. A variation on this procedure, which employs two primers that initiate extension in opposite directions from the known sequence, is described in WIPO International Patent Application

15   No.: WO 96/38591. Another such technique is known as "rapid amplification of cDNA ends" or RACE. This technique involves the use of an internal primer and an external primer, which hybridizes to a polyA region or vector sequence, to identify sequences that are 5' and 3' of a known sequence. Additional techniques include capture PCR (Lagerstrom et al., PCR Methods Applic. 1:111-19, 1991) and

20   walking PCR (Parker et al., Nucl. Acids. Res. 19:3055-60, 1991). Other methods employing amplification may also be employed to obtain a full length cDNA sequence.

Another method for amplification is the ligase chain reaction (referred to as LCR), disclosed in Eur. Pat. Appl. Publ. No. 320,308. In LCR, two

25   complementary probe pairs are prepared, and in the presence of the target sequence, each pair will bind to opposite complementary strands of the target such that they abut. In the presence of a ligase, the two probe pairs will link to form a single unit. By temperature cycling, as in PCR™, bound ligated units dissociate from the target and then serve as "target sequences" for ligation of excess probe

25

pairs. U.S. Patent No. 4,883,750, describes an alternative method of amplification similar to LCR for binding probe pairs to a target sequence.

Qbeta Replicase, described in PCT Intl. Pat. Appl. Publ. No. PCT/US87/00880, may also be used as still another amplification method in the
5   present invention. In this method, a replicative sequence of RNA that has a region complementary to that of a target is added to a sample in the presence of an RNA polymerase. The polymerase will copy the replicative sequence that can then be detected.

An isothermal amplification method, in which restriction
10 endonucleases and ligases are used to achieve the amplification of target molecules that contain nucleotide 5'-[α-thio]triphosphates in one strand of a restriction site (Walker *et al.*, 1992), may also be useful in the amplification of nucleic acids in the present invention.

Strand Displacement Amplification (SDA) is another method of
15 carrying out isothermal amplification of nucleic acids which involves multiple rounds of strand displacement and synthesis, *i.e.*, nick translation. A similar method, called Repair Chain Reaction (RCR) is another method of amplification which may be useful in the present invention and is involves annealing several probes throughout a region targeted for amplification, followed by a repair reaction
20 in which only two of the four bases are present. The other two bases can be added as biotinylated derivatives for easy detection. A similar approach is used in SDA.

Sequences can also be detected using a cyclic probe reaction (CPR). In CPR, a probe having a 3' and 5' sequences of non-target DNA and an
25 internal or "middle" sequence of the target protein specific RNA is hybridized to DNA which is present in a sample. Upon hybridization, the reaction is treated with RNaseH, and the products of the probe are identified as distinctive products by generating a signal that is released after digestion. The original template is annealed to another cycling probe and the reaction is repeated. Thus, CPR

26

involves amplifying a signal generated by hybridization of a probe to a target gene specific expressed nucleic acid.

Still other amplification methods described in Great Britain Pat. Appl. No. 2 202 328, and in PCT Intl. Pat. Appl. Publ. No. PCT/US89/01025, may be used in accordance with the present invention. In the former application, "modified" primers are used in a PCR-like, template and enzyme dependent synthesis. The primers may be modified by labeling with a capture moiety (*e.g.,* biotin) and/or a detector moiety (*e.g.,* enzyme). In the latter application, an excess of labeled probes is added to a sample. In the presence of the target sequence, the probe binds and is cleaved catalytically. After cleavage, the target sequence is released intact to be bound by excess probe. Cleavage of the labeled probe signals the presence of the target sequence.

Other nucleic acid amplification procedures include transcription-based amplification systems (TAS) (Kwoh *et al.,* 1989; PCT Intl. Pat. Appl. Publ. No. WO 88/10315), including nucleic acid sequence based amplification (NASBA) and 3SR. In NASBA, the nucleic acids can be prepared for amplification by standard phenol/chloroform extraction, heat denaturation of a sample, treatment with lysis buffer and minispin columns for isolation of DNA and RNA or guanidinium chloride extraction of RNA. These amplification techniques involve annealing a primer that has sequences specific to the target sequence. Following polymerization, DNA/RNA hybrids are digested with RNase H while double stranded DNA molecules are heat-denatured again. In either case the single stranded DNA is made fully double stranded by addition of second target-specific primer, followed by polymerization. The double stranded DNA molecules are then multiply transcribed by a polymerase such as T7 or SP6. In an isothermal cyclic reaction, the RNAs are reverse transcribed into DNA, and transcribed once again with a polymerase such as T7 or SP6. The resulting products, whether truncated or complete, indicate target-specific sequences.

Eur. Pat. Appl. Publ. No. 329,822, disclose a nucleic acid amplification process involving cyclically synthesizing single-stranded RNA ("ssRNA"), ssDNA, and double-stranded DNA (dsDNA), which may be used in accordance with the present invention. The ssRNA is a first template for a first

5   primer oligonucleotide, which is elongated by reverse transcriptase (RNA-dependent DNA polymerase). The RNA is then removed from resulting DNA:RNA duplex by the action of ribonuclease H (RNase H, an RNase specific for RNA in a duplex with either DNA or RNA). The resultant ssDNA is a second template for a second primer, which also includes the sequences of an RNA

10 polymerase promoter (exemplified by T7 RNA polymerase) 5' to its homology to its template. This primer is then extended by DNA polymerase (exemplified by the large "Klenow" fragment of *E. coli* DNA polymerase I), resulting as a double-stranded DNA ("dsDNA") molecule, having a sequence identical to that of the original RNA between the primers and having additionally, at one end, a

15 promoter sequence. This promoter sequence can be used by the appropriate RNA polymerase to make many RNA copies of the DNA. These copies can then re-enter the cycle leading to very swift amplification. With proper choice of enzymes, this amplification can be done isothermally without addition of enzymes at each cycle. Because of the cyclical nature of this process, the starting

20 sequence can be chosen to be in the form of either DNA or RNA.

PCT Intl. Pat. Appl. Publ. No. WO 89/06700, disclose a nucleic acid sequence amplification scheme based on the hybridization of a promoter/primer sequence to a target single-stranded DNA ("ssDNA") followed by transcription of many RNA copies of the sequence. This scheme is not cyclic; *i.e.,* new templates

25 are not produced from the resultant RNA transcripts. Other amplification methods include "RACE" (Frohman, 1990), and "one-sided PCR" (Ohara, 1989) which are well-known to those of skill in the art.

28

Compositions and Kits for the Detection of Cancer

The present invention further provides kits for use within any of the above diagnostic methods. Such kits typically comprise two or more components necessary for performing a diagnostic assay. Components may be compounds,

5   reagents, containers and/or equipment. For example, one container within a kit may contain a monoclonal antibody or fragment thereof that specifically binds to a lung tumor protein. Such antibodies or fragments may be provided attached to a support material, as described above. One or more additional containers may enclose elements, such as reagents or buffers, to be used in the assay. Such kits

10   may also, or alternatively, contain a detection reagent as described above that contains a reporter group suitable for direct or indirect detection of antibody binding.

The present invention also provides kits that are suitable for performing the detection methods of the present invention. Exemplary kits

15   comprise oligonucleotide primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Within certain embodiments, kits according to the present invention may also comprise a nucleic acid polymerase and suitable buffer. Exemplary oligonucleotide primers suitable for kits of the present invention are disclosed herein. Exemplary polynucleotides suitable for kits of the present

20   invention are disclosed herein.

Alternatively, a kit may be designed to detect the level of mRNA encoding a lung tumor protein in a biological sample. Such kits generally comprise at least one oligonucleotide probe or primer, as described above, that hybridizes to a polynucleotide encoding a lung tumor protein. Such an oligonucleotide may be

25   used, for example, within a PCR or hybridization assay. Additional components that may be present within such kits include a second oligonucleotide and/or a diagnostic reagent or container to facilitate the detection of a polynucleotide encoding a lung tumor protein.

29

In other related aspects, the present invention further provides compositions useful in the methods disclosed herein. Exemplary compositions comprise two or more oligonucleotide primer pairs each one of which specifically hybridizes to a distinct polynucleotide. Exemplary oligonucleotide primers suitable

5   for compositions of the present invention are disclosed herein. Exemplary polynucleotides suitable for compositions of the present invention are disclosed herein.

The following Example is offered by way of illustration and not by way of limitation.

10

EXAMPLE

MULTIPLEX DETECTION OF LUNG TUMORS

A Multiplex Real-time PCR assay was established in order to simultaneously detect the expression of four lung cancer-specific genes: L762

15   (SEQ ID NO:1), L984 (SEQ ID NO:3), L550 (SEQ ID NO:5) and L552 (SEQ ID NO:7). In contrast to detection approaches relying on expression analysis of single lung cancer-specific genes, this Multiplex assay was able to detect all lung tumor samples tested and analyze their combined mRNA expression profile in adenocarcinoma, squamous, small cell and large cell lung tumors. L552S and

20   L550S complement each other in detecting predominantly adenocarcinomas, L762S detects squamous cell carcinomas and L984P detects small cell carcinomas (see Table 1).

The primers and probes were designed to be intron spanning (exon specific) to eliminate any reactivity with genomic DNA making them suitable for

25   use in blood samples without having to DNAse treat mRNA samples. They were also designed to produce amlicons of different sizes to allow gel differentiation of end products if necessary.

The assay was carried out as follows: L552S (SEQ ID NO: 7), L550 (SEQ ID NO: 5), L762 (SEQ ID NO: 1), L984 (SEQ ID NO: 3) and specific primers,

30

and specific Taqman probes, were used to analyze their combined mRNA expression profile in lung tumors. The primers and probes are shown below:

L552S: Forward Primer (SEQ ID NO:9): 5' GACGGCATGAGCGACACACA. Reverse Primer (SEQ ID NO:10): 5' CCATGTCGCGCACTGGGATC. Probe (SEQ ID NO:11) (FAM-5' – 3'-TAMRA): CTGAAAGTCGGGATCCTACACCTGGGCA.

L550P: Forward Primer (SEQ ID NO:12): 5' GGCCACCGTCTGGATTCTTC. Reverse Primer (SEQ ID NO:13): 5' gaagaatccagacggtggcc. Probe (SEQ ID NO:14) (FAM-5' – 3'-TAMRA): CCGCCCCAAG ATCAAATCCA CAAACC.

L762S: Forward Primer (SEQ ID NO:15): 5' atggcagaggctgacagactc. Reverse Primer (SEQ ID NO:16): 5' TTCAACCACCTCAAATCCTTTCTTA. Probe (SEQ ID NO:17) (FAM-5' – 3'-TAMRA) TCGACAGCAAAGGAGAGATCAGAGCCC.

L984P: Forward Primer (SEQ ID NO:18): 5' TTACGACCCGCTCAGCCC. Reverse Primer (SEQ ID NO:19): 5' ctcccaacgccactgacaa. Probe (SEQ ID NO:20) (FAM-5'- 3'-TAMRA): CCAGGCCGAGCCCCTCAGAACC.

The assay conditions were:

Taqman protocol (7700 Perkin Elmer):

In 25 µl final volume: 1x Buffer A, 5mM MgCl, 0.2 mM dCTP, 0.2 mM dATP, 0.4 mM dUTP, 0.2 mM dGTP, 0.01 U/µl AmpErase UNG, 0.0375 U/µl TaqGold, 8% (v/v) Glycerol, 0.05% (v/v) (Sigma), Gelatin, 0.05% (v/v) (Sigma), Tween20 0.1% v/v (Sigma), 300mM of each forward and reverse primer for L762P, 50mM of each forward and reverse primer from (L552S, L984P, L550S, L984P) 2 pmol of each gene specific Taqman probe (L552S, L550S, L984P) and template cDNA. The PCR reaction was carried out at one cycle at 95OC for 10 minutes, followed by 50 cycles at 95OC for 15 seconds, 60OC for 1 minute, and

31

Since each primer set in the multiplex assay results in a band of unique length, expression signals of the four genes of interest was measured

5 individually by agarose gel analysis. The combined expression signal of all four genes can also be measured in real-time on an ABI 7700 Prism sequence detection system (PE Biosystems, Foster City, CA). Although specific primers have been described herein, different primer sequences, different primer or probe labeling and different detection systems could be used to perform this multiplex

10 assay. For example, a second fluorogenic reporter dye could be incorporated for parallel detection of a reference gene by real-time PCR. Or, for example a SYBR Green detection system could be used instead of the Taqman probe approach. Table 2 shows the reactivity of the multiplex PCR with different lung tumor types and normal lung tissue.

15

TABLE 2 Expression of Lung Cancer Multiplex Genes (L762P, L552S, L550S, L984P) in Lung Tumor and Normal Lung

Lung Tumor Type	Positive Samples/Samples Tested
Adenocarcinoma	21/24
Squamous	17/18
Large Cell	5*/5
Small Cell	5/6
Normal Lung Tissue	0/12
Total Tumors	48/53
% Positive Tumors	90.57%

Cut-off Value = Mean normal lung +3 SD =0.901
20 * One sample at cut-off

From the foregoing it will be appreciated that, although specific embodiments of the invention have been described herein for purposes of

illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited except as by the appended claims.

CLAIMS .

We Claim:

1.   A method for detecting the presence of a cancer cell in a patient, said method comprising the steps of:

    (a)   obtaining a biological sample from said patient;

    (b)   contacting the biological sample with two or more oligonucleotide pairs specific for independent polynucleotide sequences which are unrelated to one another, wherein the oligonucleotide pairs hybridize, under moderately stringent conditions, to their respective polynucleotides and the complements thereof;

    (c)   amplifying said polynucleotides; and

    (d)   detecting said amplified polynucleotides;

    wherein the presence of one or more of said amplified polynucleotides indicates the presence of lung cancer cells in said patient.

2.   A method for determining the presence of lung cancer cells in a patient, said method comprising the steps of:

    (a)   obtaining a biological sample from said patient;

    (b)   contacting a biological sample obtained from the patient with two or more oligonucleotides that hybridize to two or more polynucleotides that encode two or more lung tumor proteins;

    (c)   detecting in said biological sample an amount of a polynucleotide that hybridizes to at least one of said oligonucleotides; and

    (d)   comparing the amount of the polynucleotides that hybridizes to said oligonucleotides to a predetermined cut-off value, and therefrom determining the presence or absence of lung cancer cells in the patient.

34

3.   A method for monitoring the progression of lung cancer in a patient, said method comprising the steps of:

(a)   obtaining a first biological sample from said patient;

(b)   contacting said first biological sample with one or more oligonucleotides that hybridize to one or more polynucleotides that encode lung tumor proteins;

(c)   detecting in said first biological sample an amount of at least one of said polynucleotides that hybridize to said oligonucleotides;

(d)   repeating steps (b) and (c) using a second biological sample obtained from said patient at a subsequent point in time; and

(e)   comparing the amount of polynucleotides detected in step (d) with the amount detected in step (c) and therefrom monitoring the progression of lung cancer in said patient.

ABSTRACT OF THE DISCLOSURE

Compositions and methods for the diagnosis of lung cancer are disclosed. Such methods are useful to detect early tumors or provide adequate stage/grade information or tumor specificity. Compositions may comprise one or more lung tumor proteins, immunogenic portions thereof, or polynucleotides that encode such portions. Such compositions may be used, for example, to improve lung cancer diagnosis and prognosis and potentially differentiate between NSCLC and SCLC.

367567

SEQUENCE LISTING

<110> Zehentner, Barbara
       Hayes, Dawn
       Houghton, Raymond L.

<120> METHODS, COMPOSITIONS AND KITS FOR THE DETECTION
      AND MONITORING OF LUNG CANCER

<130> 210121.609P1

<140> US
<141> 2003-03-24

<160> 20

<170> Corixa Invention Disclosure Database

<210> 1
<211> 3951
<212> DNA
<213> Homo sapiens

<400> 1
tctgcatcca tattgaaaac ctgacacaat gtatgcagca ggctcagtgt gagtgaactg 60
gaggcttctc tacaacatga cccaaaggag cattgcaggt cctatttgca acctgaagtt 120
tgtgactctc ctggttgcct taagttcaga actcccattc ctgggagctg gagtacagct 180
tcaagacaat gggtataatg gattgctcat tgcaattaat cctcaggtac ctgagaatca 240
gaacctcatc tcaaacatta aggaaatgat aactgaagct tcattttacc tatttaatgc 300
taccaagaga agagtatttt tcagaaatat aaagatttta atacctgcca catggaaagc 360
taataataac agcaaaataa aacaagaatc atatgaaaag gcaaatgtca tagtgactga 420
ctggtatggg gcacatggag atgatccata caccctacaa tacagagggt gtggaaaaga 480
gggaaaatac attcatttca cacctaattt cctactgaat gataacttaa cagctggcta 540
cggatcacga ggccgagtgt ttgtccatga atgggcccac ctccgttggg gtgtgttcga 600
tgagtataac aatgacaaac ctttctacat aaatgggcaa aatcaaatta aagtgacaag 660
gtgttcatct gacatcacag gcatttttgt gtgtgaaaaa ggtccttgcc cccaagaaaa 720
ctgtattatt agtaagcttt ttaaagaagg atgcaccttt atctacaata gcacccaaaa 780
tgcaactgca tcaataatgt tcatgcaaag tttatcttct gtggttgaat tttgtaatgc 840
aagtacccac aaccaagaag caccaaacct acagaaccag atgtgcagcc tcagaagtgc 900
atgggatgta atcacagact ctgctgactt tcaccacagc tttcccatga cgggactga 960
gcttccacct cctcccacat tctcgcttgt agaggctggt gacaaagtgg tctgtttagt 1020
gctggatgtg tccagcaaga tggcagaggc tgacagactc cttcaactac aacaagccgc 1080
agaatttttat ttgatgcaga ttgttgaaat tcataccttc gtgggcattg ccagtttcga 1140
cagcaaagga gagatcagag cccagctaca ccaaattaac agcaatgatg atcgaaagtt 1200
gctggtttca tatctgccca ccactgtatc agctaaaaca gacatcagca tttgttcagg 1260
gcttaagaaa ggatttgagg tggttgaaaa actgaatgga aaagcttatg gctctgtgat 1320
gatattagtg accagcggag atgataagct tcttggcaat tgcttaccca ctgtgctcag 1380
cagtggttca acaattcact ccattgccct gggttcatct gcagccccaa atctggagga 1440
attatcacgt cttacaggag gtttaaagtt ctttgttcca gatatatcaa actccaatag 1500
catgattgat gctttcagta gaatttcctc tggaactgga gacattttcc agcaacatat 1560
tcagcttgaa agtacaggtg aaaatgtcaa acctcaccat caattgaaaa acacagtgac 1620
tgtggataat actgtgggca acgacactat gtttctagtt acgtggcagg ccagtggtcc 1680
tcctgagatt atattatttg atcctgatgg acgaaaatac tacacaaata tttttatcac 1740
caatctaact tttcggacag ctagtctttg gattccagga acagctaagc ctgggcactg 1800

```
gacttacacc ctgaacaata cccatcattc tctgcaagcc ctgaaagtga cagtgacctc 1860
tcgcgcctcc aactcagctg tgcccccagc cactgtggaa gcctttgtgg aaagagacag 1920
cctccatttt cctcatcctg tgatgattta tgccaatgtg aaacagggat tttatcccat 1980
tcttaatgcc actgtcactg ccacagttga gccagagact ggagatcctg ttacgctgag 2040
actccttgat gatggagcag gtgctgatgt tataaaaaat gatggaattt actcgaggta 2100
ttttttctcc tttgctgcaa atggtagata tagcttgaaa gtgcatgtca atcactctcc 2160
cagcataagc accccagccc actctattcc agggagtcat gctatgtatg taccaggtta 2220
cacagcaaac ggtaatattc agatgaatgc tccaaggaaa tcagtaggca gaaatgagga 2280
ggagcgaaag tggggcttta gccgagtcag ctcaggaggc tccttttcag tgctgggagt 2340
tccagctggc ccccaccctg atgtgtttcc accatgcaaa attattgacc tggaagctgt 2400
aaaagtagaa gaggaattga ccctatcttg dacagcacct ggagaagact ttgatcaggg 2460
ccaggctaca agctatgaaa taagaatgag taaaagtcta cagaatatcc aagatgactt 2520
taacaatgct attttagtaa atacatcaaa gcgaaatcct cagcaagctg gcatcaggga 2580
gatatttacg ttctcacccc aaatttccac gaatggacct gaacatcagc caaatggaga 2640
aacacatgaa agccacagaa tttatgttgc aatacgagca atggatagga actccttaca 2700
gtctgctgta tctaacattg cccaggcgcc tctgtttatt cccgccaatt ctgatcctgt 2760
acctgccaga gattatctta tattgaaagg agttttaaca gcaatgggtt tgataggaat 2820
catttgcctt attatagttg tgacacatca tactttaagc aggaaaaaga gagcagacaa 2880
gaaagagaat ggaacaaaat tattataaat aaatatccaa agtgtcttcc ttcttagata 2940
taagacccat ggccttcgac tacaaaaaca tactaacaaa gtcaaattaa catcaaaact 3000
gtattaaaat gcattgagtt tttgtacaat acagataaga ttttacatg gtagatcaac 3060
aaattctttt tgggggtaga ttagaaaacc cttacacttt ggctatgaac aaataataaa 3120
aattattctt taaagtaatg tctttaaagg caaagggaag ggtaaagtcg gaccagtgtc 3180
aaggaaagtt tgttttattg aggtggaaaa atagccccaa gcagagaaaa ggagggtagg 3240
tctgcattat aactgtctgt gtgaagcaat catttagtta ctttgattaa ttttttcttt 3300
ctccttatct gtgcagaaca ggttgcttgt ttacaactga agatcatgct atatttcata 3360
tatgaagccc ctaatgcaaa gctctttacc tcttgctatt ttgttatata tattacagat 3420
gaaatctcac tgctaatgct cagagatctt ttttcactgt aagaggtaac ctttaacaat 3480
atgggtatta cctttgtctc ttcataccgg ttttatgaca aaggtctatt gaatttattt 3540
gtttgtaagt ttctactccc atcaaagcag ctttttaagt tattgccttg gttattatgg 3600
atgatagtta tagcccttat aatgccttaa ctaaggaaga aaagatgtta ttctgagttt 3660
gttttaatac atatatgaac atatagtttt attcaattaa accaaagaag aggtcagcag 3720
ggagatacta accttggaa atgattagct ggctctgttt tttggttaaa taagagtctt 3780
taatcctttc tccatcaaga gttacttacc aagggcaggg gaaggggggat atagaggtcc 3840
caaggaaata aaaatcatct ttcatcttta attttactcc ttcctcttat tttttttaaaa 3900
gattatcgaa caataaaatc atttgccttt ttaattaaaa acataaaaaa a        3951
```

<210> 2
<211> 943
<212> PRT
<213> Homo sapiens

<400> 2

```
Met Thr Gln Arg Ser Ile Ala Gly Pro Ile Cys Asn Leu Lys Phe Val
1               5                   10                  15
Thr Leu Leu Val Ala Leu Ser Ser Glu Leu Pro Phe Leu Gly Ala Gly
            20                  25                  30
Val Gln Leu Gln Asp Asn Gly Tyr Asn Gly Leu Leu Ile Ala Ile Asn
        35                  40                  45
Pro Gln Val Pro Glu Asn Gln Asn Leu Ile Ser Asn Ile Lys Glu Met
    50                  55                  60
Ile Thr Glu Ala Ser Phe Tyr Leu Phe Asn Ala Thr Lys Arg Arg Val
65                  70                  75                  80
Phe Phe Arg Asn Ile Lys Ile Leu Ile Pro Ala Thr Trp Lys Ala Asn
```

```
                     85                    90                    95
Asn Asn Ser Lys Ile Lys Gln Glu Ser Tyr Glu Lys Ala Asn Val Ile
            100                   105                   110
Val Thr Asp Trp Tyr Gly Ala His Gly Asp Asp Pro Tyr Thr Leu Gln
            115                   120                   125
Tyr Arg Gly Cys Gly Lys Glu Gly Lys Tyr Ile His Phe Thr Pro Asn
        130                   135                   140
Phe Leu Leu Asn Asp Asn Leu Thr Ala Gly Tyr Gly Ser Arg Gly Arg
145                   150                   155                   160
Val Phe Val His Glu Trp Ala His Leu Arg Trp Gly Val Phe Asp Glu
                165                   170                   175
Tyr Asn Asn Asp Lys Pro Phe Tyr Ile Asn Gly Gln Asn Gln Ile Lys
            180                   185                   190
Val Thr Arg Cys Ser Ser Asp Ile Thr Gly Ile Phe Val Cys Glu Lys
        195                   200                   205
Gly Pro Cys Pro Gln Glu Asn Cys Ile Ile Ser Lys Leu Phe Lys Glu
    210                   215                   220
Gly Cys Thr Phe Ile Tyr Asn Ser Thr Gln Asn Ala Thr Ala Ser Ile
225                   230                   235                   240
Met Phe Met Gln Ser Leu Ser Ser Val Val Glu Phe Cys Asn Ala Ser
                245                   250                   255
Thr His Asn Gln Glu Ala Pro Asn Leu Gln Asn Gln Met Cys Ser Leu
                260                   265                   270
Arg Ser Ala Trp Asp Val Ile Thr Asp Ser Ala Asp Phe His His Ser
            275                   280                   285
Phe Pro Met Asn Gly Thr Glu Leu Pro Pro Pro Pro Thr Phe Ser Leu
    290                   295                   300
Val Glu Ala Gly Asp Lys Val Val Cys Leu Val Leu Asp Val Ser Ser
305                   310                   315                   320
Lys Met Ala Glu Ala Asp Arg Leu Leu Gln Leu Gln Gln Ala Ala Glu
                325                   330                   335
Phe Tyr Leu Met Gln Ile Val Glu Ile His Thr Phe Val Gly Ile Ala
            340                   345                   350
Ser Phe Asp Ser Lys Gly Glu Ile Arg Ala Gln Leu His Gln Ile Asn
        355                   360                   365
Ser Asn Asp Asp Arg Lys Leu Leu Val Ser Tyr Leu Pro Thr Thr Val
    370                   375                   380
Ser Ala Lys Thr Asp Ile Ser Ile Cys Ser Gly Leu Lys Lys Gly Phe
385                   390                   395                   400
Glu Val Val Glu Lys Leu Asn Gly Lys Ala Tyr Gly Ser Val Met Ile
            405                   410                   415
Leu Val Thr Ser Gly Asp Asp Lys Leu Leu Gly Asn Cys Leu Pro Thr
            420                   425                   430
Val Leu Ser Ser Gly Ser Thr Ile His Ser Ile Ala Leu Gly Ser Ser
            435                   440                   445
Ala Ala Pro Asn Leu Glu Glu Leu Ser Arg Leu Thr Gly Gly Leu Lys
    450                   455                   460
Phe Phe Val Pro Asp Ile Ser Asn Ser Asn Ser Met Ile Asp Ala Phe
465                   470                   475                   480
Ser Arg Ile Ser Ser Gly Thr Gly Asp Ile Phe Gln Gln His Ile Gln
            485                   490                   495
Leu Glu Ser Thr Gly Glu Asn Val Lys Pro His His Gln Leu Lys Asn
            500                   505                   510
Thr Val Thr Val Asp Asn Thr Val Gly Asn Asp Thr Met Phe Leu Val
```

```
            515                    520                    525
Thr Trp Gln Ala Ser Gly Pro Pro Glu Ile Ile Leu Phe Asp Pro Asp
    530                 535                 540
Gly Arg Lys Tyr Tyr Thr Asn Asn Phe Ile Thr Asn Leu Thr Phe Arg
545                 550                 555                 560
Thr Ala Ser Leu Trp Ile Pro Gly Thr Ala Lys Pro Gly His Trp Thr
            565                 570                 575
Tyr Thr Leu Asn Asn Thr His His Ser Leu Gln Ala Leu Lys Val Thr
            580                 585                 590
Val Thr Ser Arg Ala Ser Asn Ser Ala Val Pro Pro Ala Thr Val Glu
            595                 600                 605
Ala Phe Val Glu Arg Asp Ser Leu His Phe Pro His Pro Val Met Ile
    610                 615                 620
Tyr Ala Asn Val Lys Gln Gly Phe Tyr Pro Ile Leu Asn Ala Thr Val
625                 630                 635                 640
Thr Ala Thr Val Glu Pro Glu Thr Gly Asp Pro Val Thr Leu Arg Leu
            645                 650                 655
Leu Asp Asp Gly Ala Gly Ala Asp Val Ile Lys Asn Asp Gly Ile Tyr
            660                 665                 670
Ser Arg Tyr Phe Phe Ser Phe Ala Ala Asn Gly Arg Tyr Ser Leu Lys
            675                 680                 685
Val His Val Asn His Ser Pro Ser Ile Ser Thr Pro Ala His Ser Ile
    690                 695                 700
Pro Gly Ser His Ala Met Tyr Val Pro Gly Tyr Thr Ala Asn Gly Asn
705                 710                 715                 720
Ile Gln Met Asn Ala Pro Arg Lys Ser Val Gly Arg Asn Glu Glu Glu
            725                 730                 735
Arg Lys Trp Gly Phe Ser Arg Val Ser Ser Gly Gly Ser Phe Ser Val
            740                 745                 750
Leu Gly Val Pro Ala Gly Pro His Pro Asp Val Phe Pro Pro Cys Lys
    755                 760                 765
Ile Ile Asp Leu Glu Ala Val Lys Val Glu Glu Glu Leu Thr Leu Ser
    770                 775                 780
Trp Thr Ala Pro Gly Glu Asp Phe Asp Gln Gly Gln Ala Thr Ser Tyr
785                 790                 795                 800
Glu Ile Arg Met Ser Lys Ser Leu Gln Asn Ile Gln Asp Asp Phe Asn
            805                 810                 815
Asn Ala Ile Leu Val Asn Thr Ser Lys Arg Asn Pro Gln Gln Ala Gly
            820                 825                 830
Ile Arg Glu Ile Phe Thr Phe Ser Pro Gln Ile Ser Thr Asn Gly Pro
    835                 840                 845
Glu His Gln Pro Asn Gly Glu Thr His Glu Ser His Arg Ile Tyr Val
850                 855                 860
Ala Ile Arg Ala Met Asp Arg Asn Ser Leu Gln Ser Ala Val Ser Asn
865                 870                 875                 880
Ile Ala Gln Ala Pro Leu Phe Ile Pro Pro Asn Ser Asp Pro Val Pro
            885                 890                 895
Ala Arg Asp Tyr Leu Ile Leu Lys Gly Val Leu Thr Ala Met Gly Leu
            900                 905                 910
Ile Gly Ile Ile Cys Leu Ile Ile Val Val Thr His His Thr Leu Ser
    915                 920                 925
Arg Lys Lys Arg Ala Asp Lys Lys Glu Asn Gly Thr Lys Leu Leu
    930                 935                 940
```

<210> 3
<211> 785
<212> DNA
<213> Homo sapiens

<400> 3
```
tctgattccg cgactccttg gccgccgctg cgcatggaaa gctctgccaa gatggagagc  60
ggcggcgccg gccagcagcc ccagccgcag ccccagcagc ccttcctgcc gcccgcagcc 120
tgtttctttg ccacggccgc agccgcggcg gccgcagccg ccgcagcggc agcgcagagc 180
gcgcagcagc agcagcagca gcagcagcag caggcgccgc agctgagacc ggcggccgac 240
ggccagccct caggggggcgg tcacaagtca gcgcccaagc aagtcaagcg acagcgctcg 300
tcttcgcccg aactgatgcg ctgcaaacgc cggctcaact tcagcggctt tggctacagc 360
ctgccgcagc agcagccggc cgccgtggcg cgccgcaacg agcgcgagcg caaccgcgtc 420
aagttggtca acctgggctt tgccacccct cgggagcacg tccccaacgg cgcggccaac 480
aagaagatga gtaaggtgga gacactgcgc tcggcggtcg agtacatccg cgcgctgcag 540
cagctgctgg acgagcatga cgcggtgagc gccgccttcc aggcaggcgt cctgtcgccc 600
accatctccc ccaactactc caacgacttg aactccatgg ccggctcgcc ggtctcatcc 660
tactcgtcgg acgagggctc ttacgacccg ctcagccccg aggagcagga gcttctcgac 720
ttcaccaact ggttctgagg ggctcggcct ggtcaggccc tggtgcgaat ggactttgga 780
agcag                                                             785
```

<210> 4
<211> 236
<212> PRT
<213> Homo sapiens

<400> 4
```
Met Glu Ser Ser Ala Lys Met Glu Ser Gly Gly Ala Gly Gln Gln Pro
1               5                  10                  15
Gln Pro Gln Pro Gln Gln Pro Phe Leu Pro Pro Ala Ala Cys Phe Phe
            20                  25                  30
Ala Thr Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Ala Gln
        35                  40                  45
Ser Ala Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Ala Pro
    50                  55                  60
Gln Leu Arg Pro Ala Ala Asp Gly Gln Pro Ser Gly Gly Gly His Lys
65                  70                  75                  80
Ser Ala Pro Lys Gln Val Lys Arg Gln Arg Ser Ser Ser Pro Glu Leu
                85                  90                  95
Met Arg Cys Lys Arg Arg Leu Asn Phe Ser Gly Phe Gly Tyr Ser Leu
            100                 105                 110
Pro Gln Gln Gln Pro Ala Ala Val Ala Arg Arg Asn Glu Arg Glu Arg
        115                 120                 125
Asn Arg Val Lys Leu Val Asn Leu Gly Phe Ala Thr Leu Arg Glu His
    130                 135                 140
Val Pro Asn Gly Ala Ala Asn Lys Lys Met Ser Lys Val Glu Thr Leu
145                 150                 155                 160
Arg Ser Ala Val Glu Tyr Ile Arg Ala Leu Gln Gln Leu Leu Asp Glu
                165                 170                 175
His Asp Ala Val Ser Ala Ala Phe Gln Ala Gly Val Leu Ser Pro Thr
            180                 185                 190
Ile Ser Pro Asn Tyr Ser Asn Asp Leu Asn Ser Met Ala Gly Ser Pro
        195                 200                 205
Val Ser Ser Tyr Ser Ser Asp Glu Gly Ser Tyr Asp Pro Leu Ser Pro
```

```
          210               215               220
Glu Glu Gln Glu Leu Leu Asp Phe Thr Asn Trp Phe
225                       230               235
```

<210> 5
<211> 1633
<212> DNA
<213> Homo sapiens

<400> 5

```
cgtggaggca gctagcgcga ggctggggag cgctgagccg cgcgtcgtgc cctgcgctgc 60
ccagactagc gaacaataca gtcgggatgg ctaaaggtga ccccaagaaa ccaaagggca 120
agacgtccgc ttatgccttc tttgtgcaga catgcagaga agaacataag aagaaaaacc 180
cagaggtccc tgtcaatttt gcggaatttt ccaagaagtg ctctgagagg tggaagacgg 240
tgtccgggaa agagaaatcc aaatttgatg aaatggcaaa ggcagataaa gtgcgctatg 300
atcgggaaat gaaggattat ggaccagcta agggaggcaa gaagaagaag gatcctaatg 360
ctcccaaaag gccaccgtct ggattcttcc tgttctgttc agaattccgc cccaagatca 420
aatccacaaa ccccggcatc tctattggag acgtggcaaa aaagctgggt gagatgtgga 480
ataatttaaa tgacagtgaa aagcagcctt acatcactaa ggcggcaaag ctgaaggaga 540
agtatgagaa ggatgttgct gactataagt cgaaggaaa gtttgatggt gcaaagggtc 600
ctgctaaagt tgcccggaaa aaggtggaag aggaagatga agaacaggag gaggaagaag 660
aggaggagga ggaggaggag gatgaataaa gaaactgttt atctgtctcc ttgtgaatac 720
ttagagtagg ggagcgccgt aattgacaca tctcttattt gagaagtgtc tgttgccctc 780
attaggttta attacaaaat ttgatcacga tcatattgta gtctctcaaa gtgctctaga 840.
aattgtcagt ggtttacatg aagtggccat gggtgtctgg agcaccctga aactgtatca 900
aagttgtaca tatttccaaa catttttaaa atgaaaaggc actctcgtgt tctcctcact 960
ctgtgcactt tgctgttggt gtgacaaggc atttaaagat gtttctggca ttttcttttt 1020
atttgtaagg tggtggtaac tatggttatt ggctagaaat cctgagtttt caactgtata 1080
tatctatagt ttgtaaaaag aacaaaacaa ccgagacaaa cccttgatgc tccttgctcg 1140
gcgttgaggc tgtgggggaag atgccttttg ggagaggctg tagctcaggg cgtgcactgt 1200
gaggctggac ctgttgactc tgcagggggc atccatttag cttcaggttg tcttgtttct 1260
gtatatagtg acatagcatt ctgctgccat cttagctgtg acaaagggg ggtcagctgg 1320
catgagaata tttttttta agtgcggtag tttttaaact gtttgttttt aaacaaacta 1380
tagaactctt cattgtcagc aaagcaaaga gtcactgcat caatgaaagt tcaagaacct 1440
cctgtactta aacacgattc gcaacgttct gttattttt ttgtatgttt agaatgctga 1500
aatgttttg aagttaaata aacagtatta catttttaga actcttctct actataacag 1560
tcaatttctg actcacagca gtgaacaaac ccccactccg ttgtatttgg agactggcct 1620
ccctataaat gtg                                                   1633
```

<210> 6
<211> 200
<212> PRT
<213> Homo sapiens

<400> 6

```
Met Ala Lys Gly Asp Pro Lys Lys Pro Lys Gly Lys Met Ser Ala Tyr
1               5               10              15
Ala Phe Phe Val Gln Thr Cys Arg Glu Glu His Lys Lys Lys Asn Pro
            20              25              30
Glu Val Pro Val Asn Phe Ala Glu Phe Ser Lys Lys Cys Ser Glu Arg
        35              40              45
Trp Lys Thr Met Ser Gly Lys Glu Lys Ser Lys Phe Asp Glu Met Ala
    50              55              60
Lys Ala Asp Lys Val Arg Tyr Asp Arg Glu Met Lys Asp Tyr Gly Pro
```

```
      65                    70                    75                    80
      Ala Lys Gly Gly Lys Lys Lys Lys Asp Pro Asn Ala Pro Lys Arg Pro
                      85                    90                    95
      Pro Ser Gly Phe Phe Leu Phe Cys Ser Glu Phe Arg Pro Lys Ile Lys
                      100                   105                   110
      Ser Thr Asn Pro Gly Ile Ser Ile Gly Asp Val Ala Lys Lys Leu Gly
                  115                   120                   125
      Glu Met Trp Asn Asn Leu Asn Asp Ser Glu Lys Gln Pro Tyr Ile Thr
                  130                   135                   140
      Lys Ala Ala Lys Leu Lys Glu Lys Tyr Glu Lys Asp Val Ala Asp Tyr
      145                   150                   155                   160
      Lys Ser Lys Gly Lys Phe Asp Gly Ala Lys Gly Pro Ala Lys Val Ala
                      165                   170                   175
      Arg Lys Lys Val Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Glu Glu
                  180                   185                   190
      Glu Glu Glu Glu Glu Glu Asp Glu
                  195                   200
```

<210> 7
<211> 781
<212> DNA
<213> Homo sapiens

<400> 7
```
gcggcggagc tgtgagccgg cgactcgggt ccctgaggtc tggattcttt ctccgctact 60
gagacacggc gggtaggtcc acaggcagat ccaactggga gttgaagtgt gagtgagagt 120
gaagaggaac cagcaggctt ccggagggtt gtgtggtcag tgactcagag tgagaaggcc 180
ctcgaagtcg tcgtccctct catgcggtgc cacgcccatg gaccttcttg tctcgtcacg 240
gccataacta gggaggaagg agggccgagg agtggagggg ctcaggcgaa gctggggtgc 300
tgttgggggt atccgagtcc cagaagcacc tggaaccccg acagaagatt ctggactccc 360
cagacgggac caggagaggg acggcatgag cgacacacac aaacacagaa ccacacagcc 420
agtcccagga gcccagtaat ggagagcccc aaaaagaaga accagcagct gaaagtcggg 480
atcctacacc tgggcagcag acagaagaag atcaggatac agctgagatc ccagtgcgcg 540
acatggaagg tgatctgcaa gagctgcatc agtcaaacac cggggataaa tctggatttg 600
ggttccggcg tcaaggtgaa gataatacct aaagaggaac actgtaaaat gccagaagca 660
ggtgaagagc aaccacaagt ttaaatgaag acaagctgaa acaacgcaag ctggttttat 720
attagatatt tgacttaaac tatctcaata aagttttgca gctttcacca aaaaaaaaaa 780
a                                                                   781
```

<210> 8
<211> 160
<212> PRT
<213> Homo sapiens

<400> 8
```
      Met Arg Cys His Ala His Gly Pro Ser Cys Leu Val Thr Ala Ile Thr
      1                   5                    10                   15
      Arg Glu Glu Gly Gly Pro Arg Ser Gly Gly Ala Gln Ala Lys Leu Gly
                      20                    25                   30
      Cys Cys Trp Gly Tyr Pro Ser Pro Arg Ser Thr Trp Asn Pro Asp Arg
                  35                    40                    45
      Arg Phe Trp Thr Pro Gln Thr Gly Pro Gly Glu Gly Arg His Glu Arg
          50                    55                    60
      His Thr Gln Thr Gln Asn His Thr Ala Ser Pro Arg Ser Pro Val Met
```

```
65                      70                    .75                  80
Glu Ser Pro Lys Lys Lys Asn Gln Gln Leu Lys Val Gly Ile Leu His
                85                    90                  95
Leu Gly Ser Arg Gln Lys Lys Ile Arg Ile Gln Leu Arg Ser Gln Cys
            100               105               110
Ala Thr Trp Lys Val Ile Cys Lys Ser Cys Ile Ser Gln Thr Pro Gly
        115               120               125
Ile Asn Leu Asp Leu Gly Ser Gly Val Lys Val Lys Ile Ile Pro Lys
    130               135               140
Glu Glu His Cys Lys Met Pro Glu Ala Gly Glu Glu Gln Pro Gln Val
145               150               155               160
```

<210> 9
<211> 20
<212> DNA
<213> Homo sapiens

<400> 9
gacggcatga gcgacacaca   20

<210> 10
<211> 20
<212> DNA
<213> Homo sapiens

<400> 10
ccatgtcgcg cactgggatc   20

<210> 11
<211> 28
<212> DNA
<213> Homo sapiens

<400> 11
ctgaaagtcg ggatcctaca cctgggca   28

<210> 12
<211> 20
<212> DNA
<213> Homo sapiens

<400> 12
ggccaccgtc tggattcttc   20

<210> 13
<211> 20
<212> DNA
<213> Homo sapiens

<400> 13
gaagaatcca gacggtggcc   20

<210> 14